

Dynamic Diagonal Operators, Walsh Mixing, and Low-Loss Neural Geometry

Introduction

A surprisingly large range of neural-network and signal-processing systems can be viewed through a common mathematical structure:

- a fixed mixing transform,
- combined with a data-dependent diagonal operator.

This viewpoint unifies several ideas that at first appear unrelated:

- ReLU neural-network gating,
- sign-based switching systems,
- Walsh-Hadamard transforms,
- sparse routing,
- reversible dynamics,
- projection geometry,
- binary and ternary operator systems.

The central observation is that a diagonal matrix whose entries are chosen dynamically from a small set such as:

- $\{0,1\}$
- $\{-1,1\}$
- $\{-1,0,1\}$

can generate surprisingly rich behavior when combined with strong mixing transforms such as the Fast Walsh-Hadamard Transform (FWHT).

Even more interestingly, these systems can interpolate between:

- destructive neural gating,
- reversible wave-like dynamics,
- symbolic switching systems,
- adaptive graph structures,
- and low-information-loss neural computation.

This note introduces the main ideas and some promising directions for further exploration.

1. ReLU Networks as Dynamic Projection Systems

A ReLU layer can be interpreted geometrically as a data-dependent projection.

For each input vector, ReLU decides which coordinates survive and which are suppressed. This decision can be represented by a diagonal matrix whose entries are either:

- 1 for active neurons,
- 0 for inactive neurons.

The diagonal matrix is idempotent, meaning that applying it twice has the same effect as applying it once. Geometrically, it projects vectors onto a coordinate subspace.

This perspective reveals several important properties:

- information is destroyed whenever coordinates are zeroed,
- the rank of the system can decrease,
- some directions in state space vanish permanently,
- adjacent matrices inherit zero rows or columns depending on multiplication order.

In other words, ReLU networks are not merely nonlinear scalar systems. They are dynamically changing projection geometries.

2. Sign-Based Diagonal Operators

A different class of systems appears when the diagonal entries are restricted to:

- -1 or +1.

Instead of deleting coordinates, these operators reflect coordinates.

Unlike ReLU gating, these matrices preserve magnitude and are reversible. Applying the operator twice returns the original vector.

This changes the character of the dynamics completely:

- energy is preserved,

- information loss is greatly reduced,
- the system behaves more like a reflection process than a projection process.

When combined with strong mixing transforms, these sign operators create highly structured but dynamically changing global interactions.

3. Walsh-Hadamard Mixing

The Fast Walsh-Hadamard Transform is particularly interesting because it provides:

- dense global mixing,
- efficient computation,
- binary structure,
- hierarchical interaction patterns.

When a diagonal switching matrix is inserted before or after a Walsh transform, the resulting operator behaves like a dynamically changing signed interaction graph.

The transform spreads local sign changes globally. A single diagonal decision can influence many coordinates after mixing.

Repeated compositions such as:

- HD,
- DHD,
- HDH,
- HD_1HD_2 ,

produce alternating stages of:

- global scrambling,
- local selection,
- redistribution.

This creates behavior reminiscent of:

- graph diffusion,
 - sparse routing,
 - butterfly networks,
 - reversible computation,
 - symbolic dynamics.
-

4. Ternary Systems: $\{-1,0,1\}$

One particularly interesting regime uses diagonal entries from:

- -1,
- 0,
- +1.

This combines three operations simultaneously:

Value	Interpretation
0	delete or suppress
+1	preserve orientation
-1	reflect orientation

These ternary systems are hybrids between:

- reversible sign systems,
- and destructive ReLU-like systems.

They allow selective information destruction while preserving reversibility elsewhere.

This may provide a useful compromise between:

- expressive nonlinear behavior,
 - and low information loss.
-

5. Data-Dependent Operator Selection

The diagonal entries can be generated in many ways.

Simple examples include:

- $\text{sign}(x)$,
- $\text{sign}(Hx)$,
- thresholding,
- ReLU selection,
- randomized projections.

The important insight is that the nonlinearity is no longer merely an activation function. Instead, it becomes a mechanism for selecting operators dynamically.

The system is effectively switching among many different linear transforms depending on the input.

This viewpoint connects naturally to:

- switched dynamical systems,
 - symbolic dynamics,
 - finite-state systems,
 - adaptive graph computation,
 - and sparse neural routing.
-

6. The Danger of $\text{abs}()$ -Type Collapse

Some selection rules can unintentionally destroy information.

For example, if the diagonal entries are chosen using $\text{sign}(x)$, then multiplying the vector by that diagonal matrix converts every coordinate to its absolute value.

Positive and negative states become identified.

Geometrically, opposite regions of space collapse together.

This is similar to how ReLU folds negative space onto positive space.

Although the system may remain highly nonlinear, orientation information is lost.

7. Spectral and Randomized Selection

Using transformed coordinates for the decisions changes the geometry significantly.

Examples include:

- $\text{sign}(Hx)$,
- $\text{sign}(RHx)$,

where R is some scrambling or random mixing matrix.

In these systems:

- decisions occur in transformed coordinate systems,
- switching boundaries become global rather than local,
- correlations are reduced,
- repeated alignment problems become less severe.

Randomized selection geometry can dramatically improve:

- state-space exploration,
- mixing quality,
- and information preservation.

The resulting systems begin to resemble:

- randomized hashing,
 - attention routing,
 - random feature systems,
 - and adaptive spectral processing.
-

8. Dynamic Graph Interpretation

These operators can also be interpreted as dynamically changing graphs.

The diagonal matrix partitions coordinates into groups:

- positive,
- negative,
- suppressed.

The surrounding transform then couples those groups globally.

The effective interaction graph changes continuously with the input data.

This creates a form of adaptive graph computation where:

- the graph structure itself depends on the signal,
 - but the computational cost remains very low.
-

9. Soft Diagonal Operators

Hard binary switching is only one possibility.

The diagonal entries can instead be continuous functions constrained to remain near useful operating regions.

Examples include:

- $\tanh(x)$,
- signed square root functions,
- normalized sign-preserving nonlinearities.

This introduces a continuum between:

- hard reflections,
 - and soft gain control.
-

10. Signed Square Root Dynamics

A particularly interesting example is the signed square root function:

- positive values map to their square root,
- negative values map to the negative square root of their magnitude.

This function has an unusual property:

- values near zero are pushed away from zero,
- while large values are compressed.

Most neural nonlinearities suppress small signals. Signed square root does the opposite.

This has several important consequences:

- weak coordinates are amplified instead of dying,
- near-degenerate directions are revived,
- information loss from tiny diagonal values is reduced,
- the system naturally drifts toward stable ± 1 states.

Over repeated application, the diagonal entries tend to stabilize near:

- $+1$,

- or -1.

This means the system gradually approaches a nearly reversible sign-based regime while remaining continuous and smooth.

11. Information Preservation

Standard neural networks often lose information rapidly because:

- ReLU destroys dimensions,
- small activations vanish,
- repeated contractions accumulate.

Sign-based and softly normalized diagonal systems behave differently.

They can preserve:

- orientation,
- energy,
- weak signal components,
- and reversible structure.

This suggests the possibility of neural architectures with much lower information loss per layer.

Such systems may allow:

- deeper stable networks,
 - reversible computation,
 - efficient iterative dynamics,
 - and improved state-space coverage.
-

12. Dynamic Diagonal Operators as Local Geometry Modifiers

The most useful conceptual shift may be this:

Diagonal nonlinearities are not merely activation functions.

They are local geometry modifiers.

The diagonal entries determine:

- which directions survive,
- which reflect,
- which expand,
- which contract,
- and how local structure interacts with global mixing.

The surrounding transform determines how these local geometric decisions propagate through the entire system.

From this perspective, many neural and signal-processing systems can be understood as alternating stages of:

- global mixing,
 - and local geometry modification.
-

13. Future Directions

Several areas appear especially promising:

Low-information-loss neural networks

Architectures that avoid unnecessary projection and contraction.

Reversible neural systems

Using sign-preserving dynamics instead of destructive gating.

Sparse adaptive routing

Dynamic operator selection using tiny discrete alphabets.

Spectral nonlinear computation

Applying gating decisions in transformed domains rather than coordinate space.

Symbolic dynamical systems

Viewing switching patterns as finite-state symbolic processes.

Hardware-efficient neural layers

Binary and ternary diagonal operators are extremely cheap computationally.

Dynamical graph systems

Interpreting transformed diagonal operators as adaptive signed graphs.

Conclusion

A diagonal matrix with dynamically chosen entries may appear mathematically simple, but when combined with strong mixing transforms it becomes a powerful mechanism for constructing adaptive geometry.

Depending on how the diagonal entries are generated and constrained, the resulting systems can behave like:

- projection systems,
- reflection systems,
- graph dynamical systems,
- reversible mixers,
- sparse routing networks,
- or low-loss neural operators.

The design space is extremely large and still relatively unexplored.

One of the most intriguing possibilities is that useful neural computation may not require increasingly complicated nonlinearities. Instead, rich behavior may emerge from the interaction between:

- strong mixing transforms,
- simple discrete operator alphabets,
- and dynamically changing local geometry.